NAVAL HEALTH RESEARCH CENTER

CONFIRMATORY FACTOR ANALYSIS OF THE CENTER FOR EPIDEMIOLOGICAL STUDIES DEPRESSION SCALE (CES-D) IN MILITARY RECRUITS

R. R. Vickers, Jr.

Report No. 92-29

NAVAL HEALTH RESEARCH CENTER
P. O. BOX 85122
SAN DIEGO, CALIFORNIA 92186 – 5122

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND

Confirmatory Factor Analysis of the Center for Epidemiological Studies Depression Scale (CES-D) in Military Recruits[†]

Ross R. Vickers, Jr.

Cognitive Performance and Psychophysiology Department
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

[†]Technical Report 92-29 supported by the Naval Medical Research and Development Command, Navy Medical Command, Bureau of Medicine and Surgery, under Research Work Unit MR04101.00A-6004 and by the Office of Naval Research under Reimbursable 6252. The views presented are those of the author and do not reflect the official policy of the Department of the Navy, Department of Defense, nor the U.S. Government.

Summary

Problem

The psychological reactions of military personnel to stress are a frequent concern for military medicine. Depression, a common psychological reaction to stress, can be measured by a number of standardized instruments. However, routine application of standardized instruments to military populations may be a questionable practice. Symptom reports in these populations may be affected by situational demands (e.g., training requirements) and differences in the way symptoms are expressed in a young, generally healthy population compared to the population at large.

Objective

The present study sought to demonstrate that a measurement model for depression developed in civilian populations could be applied to a military population which faced substantial situational challenges and significant constraints on behavior.

Approach

U.S. Navy recruits completed the Center for Epidemiological Studies Depression scale (CES-D) one week after arriving at the Recruit Training Command, San Diego. The CES-D is a 20-item scale comprised of symptoms indicating depressed mood, lack of positive mood, disruption of behavior (e.g., sleep, thought, appetite), and feelings of social isolation. Structural equation modeling was applied to compare four competing measurement models in terms of their effectiveness in reproducing the observed covariance matrices generated from the responses.

Results

A four-dimensional model developed previously in research on civilian populations fit the data as well as any of the other models considered, thereby confirming that the measurement structure for the CES-D is consistent from military to civilian populations. However, a 3-dimensional model derived from previous exploratory factor analyses in civilian populations fit the data nearly as well. Detailed examination of the misfit between the civilian 4-dimensional model and the data indicated that the best model was one which modified three factor loadings in the original civilian model.

Conclusions

The CES-D measurement model from civilian populations can be generalized to military populations with some confidence. This conclusion assumes that basic training represents a major change in the structure of daily events and that what is typical in other settings within the military typically will involve less extreme changes. Confirmatory analyses such as the present ones provide a useful check on the validity of such generalizations in specific settings. In addition, further work is desirable to choose between the 3- and 4-dimensional measurement models for depression. The major implication of the present study in this regard is that whatever refinements prove useful for representing depression in military personnel are likely to apply to the general population and vice versa. Thus, the CES-D can be used to measure depression in military personnel for comparison to other populations.

Introduction

Personnel morale is important for military units, and surveys to assess satisfaction and well-being are commonplace. These surveys frequently rely on measures of psychological status originally developed and validated in civilian populations. One advantage of this approach is that evidence of construct validity that has accumulated in the development of the measures can be invoked as a basis for claiming validity in the military survey. This generalization is legitimate provided conditions which would affect the measures are comparable in the two populations, but the generalization can be misleading if the demands of training and operational conditions in the military alter the significance of the indicators comprising the survey measure. In Meehl's (1991) terminology, the ceteris paribus (i.e., "all other things equal") condition for valid hypothesis tests would not hold in this case.

Tests of the validity of ceteris paribus assumptions are needed to determine whether it is legitimate to compare results obtained in military populations to norms developed in civilian populations. In the case of military personnel, the demands of operational or training conditions could selectively alter scores on some indicators of morale (e.g., fatigue symptoms). Comparing scores under these circumstances to civilian norms might convey a misleading impression that morale was suffering when all that was actually observed was a normal response to the demands of the situation.

Structural equation modeling provides one means of testing ceteris paribus assumptions when generalizing from civilian to military populations. If situational factors affect only some indicators in a morale measure, it would be reasonable to expect that the pattern of covariation between the affected indicators and other indicators in a morale composite would change. This assertion follows from the fact that covariations between indicators are assumed to be the product of any causal effects of one indicator on another or any common causes of variation in the different indicators (Bentler & Woodward, 1980). If ceteris paribus holds, these sources of covariation are constant; if any specific situational source(s) of variance affect several indicators, these situational factors will modify the covariance structure unless they are so delicately balanced as to produce the same structure as a product of different causes. The extent of differences in the covariation of indicators comprising morale measures, therefore, can be an index of the legitimacy of the ceteris paribus assumption. Confirmatory factor analysis provides

a method of testing for differences between covariance matrices, quantifying the magnitude of those differences, and identifying the location of the differences.

The present study applied confirmatory factor analyses to assess the legitimacy of the ceteris paribus assumption when a measure of depression is generalized from civilian to military populations. Depression is a common mental health problem (e.g., Regier et al., 1988) and would signal significant psychological impairment in military personnel. In recent years, the Center for Epidemiological Studies-Depression Scale (CES-D; Radloff, 1977) has been used with increasing frequency for research which requires measuring depression in the general United States population. Whenever a psychological instrument is widely used, it is possible that the measurements provided are not invariant across the populations and situations evaluated (Blalock, 1982). While researchers commonly assume that measurements are invariant, this assumption needs to be subjected to appropriate tests to ensure that the auxiliary measurement theory is valid when testing specific hypotheses in particular groups (Meehl, 1990; 1991). Ideally, the measurement of a given construct provided by a particular instrument will be robust across samples and situations, thereby facilitating the interpretation of results within any single study and the aggregation of results across studies. Useful information, therefore, is provided by examining the robustness of measurement models across samples with different characteristics, particularly for widely used instruments such as the CES-D.

Hertzog, Van Alstine, Usala, Hultsch, and Dixon (1990) recently applied confirmatory factor analysis to demonstrate that a single measurement model for the CES-D was suitable for two samples of subjects. One sample was comprised of residents of Annville, Pennsylvania, between 20 and 80 years of age. The other sample was comprised of residents of Victoria, British Columbia, between 55 and 78 years of age. Hertzog et al.'s (1990) finding that a single measurement model fit the data for both samples suggests that the CES-D provides a robust measurement framework across these different populations. However, the actual range of social differences sampled in that study is uncertain, leaving room for doubt about how widely the Hertzog et al. (1990) findings will generalize.

In the present study, the robustness of the CES-D measurement structure was evaluated further by studying depressive symptomatology in a specific situation in which the age distribution, situational stresses, and ecological constraints on behavior differed from those in the

Hertzog et al. (1990) study. To this end, the Hertzog et al. (1990) model was applied to data obtained from U.S. Navy recruits early in basic training. Most recruits are younger than 20 years of age, the lower limit for either Hertzog et al. (1990) samples. All recruits face adaptational challenges that are standardized with regard to general temporal sequence and objective adaptational requirements (Bourne, 1967; Zurcher, 1968). These challenges consistently elicit emotional reactions that include depressed mood (Datel, Engle, & Barba, 1966; LaRocco, Ryman, & Biersner, 1977). Relative to the recruit situation, studies of community samples such as those used by Hertzog et al. (1990) could be expected to involve less uniformity across participants regarding the types of recent or concurrent stresses experienced and the timing of the onset of those stresses relative to completion of the CES-D. Finally, recruits' eating, work, and sleeping arrangements are standardized, their opportunities for social interaction are restricted, and they are separated from friends and family. Each of these aspects of basic training could produce reports of behavior that would be indicative of depression under ordinary circumstances, but that may have substantially different significance in this special setting. Given these considerations, evidence that Hertzog et al.'s (1990) measurement model fits data from recruits would add to the empirical basis for claiming that this structure is not strongly dependent on age, on the type, magnitude, and timing of recent stresses, or on specific situational determinants of reported symptoms.

Hertzog Model

The Hertzog et al. (1990) model was comprised of four correlated dimensions labelled "depression" (e.g., feeling blue, sad), "somatic symptoms" (e.g., trouble sleeping, trouble concentrating), "loss of well-being" (e.g., less happy or hopeful than usual), and "interpersonal difficulties" (e.g., people are unfriendly). These four dimensions correspond to those shown in Table 1, but different labels have been used in this paper to facilitate discussion.

"Depression" has been labelled "Depressed Mood" to emphasize that this dimension reflects feelings and emotional states using terms that are very similar to those commonly found in mood questionnaire scales of depression (McNair, Lorr, & Droppleman, 1971; Ryman, Biersner, & LaRocco, 1974). This distinction may help avoid confusing scores on this dimension with the full clinical syndrome of depression.

"Somatic Symptoms" has been labelled "Functional Disruption." This change seemed justified on the grounds that the items which are indicators of this dimension include disruption of common activities rather than somatic changes such as weight loss or weight gain which are indicated in some depression inventories. The concept of functional disruption underlying the change in labels is consistent with the use of this term by Wolff, Hofer, and Mason (1964) as a component of effectiveness of defenses.

"Well-being" has been labelled "Positive Affect" to underscore the emphasis on positive feelings in the items comprising this dimension. The content of the items suggests that this dimension perhaps should be aligned with concepts such as positive affectivity (Watson, Clark, & Tellegen, 1988; Watson & Tellegen, 1985) and happiness (Fordyce, 1988) which are often found in the literature on emotion rather than being equated with the general concept of well-being. This suggestion is based on the observation that well-being can be defined as including, for example, measures of life satisfaction and quality of life which are not necessarily highly correlated with positive affect.

The label changes were introduced in the belief that they emphasize similarities between CES-D components and common affective concepts. No changes were made regarding the items employed as markers for the different dimensions or the correlations between dimensions. Therefore, to test the assumption of measurement invariance from the original samples to the present sample, the factor loadings for items and the correlations between factors were fixed at values determined by taking the simple weighted average of the figures for the two samples reported by Hertzog et al. (1990).

<u>Alternative Measurement Models</u>

A test of the hypothesis that the Hertzog et al. (1990) measurement model fits the data from military recruits can be conducted by confirmatory factor analysis. However, simply fitting their model to the data will not suffice. Confirmatory factor analyses must contrast multiple models, even when one model has a high probability of being correct. If no comparisons are made, only statistical criteria corresponding to standard significance tests can be applied to evaluate the model, a procedure which can be misleading when only a single model is considered. For example, if the Hertzog et al. (1990) model fit the data well by statistical criteria, it could be accepted as the most appropriate model even though other plausible models actually fit the data

better. On the other hand, the Hertzog et al. (1990) model could fit the data poorly by statistical fit criteria (e.g., many of which are sensitive to sample size), but still fit the data better than any plausible alternative model. To guard against this type of problem, five alternative measurement models were compared in this study.

The potential problems associated with evaluating the Hertzog et al. (1990) model in isolation were dealt with by considering four other plausible measurement models as well. These alternative models were:

(a) Unidimensional Model: This model assumed that the CES-D provided a unidimensional measure of depression. All items, therefore, were permitted to load on a single factor. This model corresponds to the concept of depression as a unitary construct rather than the composite of relatively independent subunits implied by the four factors in the Hertzog et al. (1990) model.

(b) Sample Weights Model: This model was derived from the Hertzog model by retaining the pattern of factor loadings and factor correlations, but estimating the values for the model parameters from the recruit data. In practice, this meant that a four-dimensional model was estimated in which each dimension was defined by the same set of items as in the Hertzog et al. (1990) model, but the size of the factor loading was not specified as it was when the Hertzog model was fitted to the data.

(c) Orthogonal Model: This model was based on published exploratory analyses for the CES-D (Berkman et al., 1986; Clark, Aneshensel, Frerichs, & Morgan, 1981; Devins et al., 1988; Ensel, 1986; Kuo, 1984; Radloff, 1977; Roberts, 1980; Ross & Mirowsky, 1984). Although many CES-D items consistently load on more than one dimension when orthogonal components are extracted and rotated (Table 1), Hertzog et al. (1990) treated each item as an indicator of a single dimension. The working assumption apparently was that the factorial complexity of the items in the exploratory analysis was a by-product of the decision to extract orthogonal components. The choice between extracting correlated dimensions to increase the approximation to simple structure (i.e., having items load on a single dimension) or orthogonal dimensions which have the desirable statistical characteristic of defining uncorrelated constructs is a fundamental decision in all factor analysis (Gorsuch, 1973). In general, correlated dimensions can be justified on the grounds that the results obtained with this approach are more robust across samples and the argument that psychological constructs generally cannot be expected to be completely independent of one another. The latter point has special force in the present case, because the four dimensions measured are components of a single higher-order construct. Despite the fundamental plausibility of the Hertzog et al. (1990) approach, it was considered worthwhile for the purposes of the present study to include a model which represented the

alternative choice with four orthogonal dimensions with multiple loadings for some items to test the legitimacy of this decision. The intent was to ensure a wide enough range of alternative substantive models to provide a suitable context for evaluating the Hertzog et al. (1990) model. Although it might be argued that this alternative was not very plausible on the grounds that the specific factors being assessed are known to be part of a well-established behavioral syndrome, its inclusion did help ensure a range of substantive alternatives that was broad enough for the present purposes.

Table 1
Average Factor Loadings for CES-D Items from Prior Studies

Item	Depressed Mood	Functional Disruption	Positive Affect	Inter-personal Problems
1 Bothered by things	.29*	.47*	.02	.06
2 Poor appetite	.17	.48*	-.03	.06
3 Have the blues	.48*	.44*	-.01	.14
4 Feel as good as ever	.04	.01	.59*	-.07
5 Can't keep mind on things	.29*	.44*	-.06	.17*
6 Feel depressed	.56*	.43*	-.01	.18
7 Everything is an effort	.18	.57*	.04	.15
8 Hopeful about the future	-.04	.05	.61*	.02
9 Feel like a failure	.35*	.17*	.03	.33
10 Fearful	.50*	.27*	.05	.19*
11 Sleep is disrupted	.20*	.54*	.03	.02
12 Feel happy	-.02	-.02	.67*	.04
13 Talk less than usual	.24	.33*	-.02	.22
14 Feel lonely	.54*	.24*	.00	.24
15 People are unfriendly	.14	.14	-.03	.68*
16 Enjoy life	-.04	-.02	.69*	.04
17 Feel like crying	.63*	.09	.00	.12
18 Sad	.67*	.30*	.00	.18*
19 People dislike me	.20*	.10	.00	.73*
20 Cannot get going	.13	.63*	.01	.19

NOTE: Entries are the weighted average component or factor loadings for the 4-component solutions reported in Berkman et al. (1986), Clark et al. (1981; male sample only), Devins et al. (1988), Ensel (1986), Kuo (1984), Radloff (1977; male and female samples), Roberts (1980; Anglo, Black, and Chicano samples), and Ross & Mirowsky (1984; male and female samples; first four factors only used for males). Weights were based on the sample sizes for prior analyses. "*" indicates that the average loading was greater than twice as large as the standard deviation across samples.

For the Orthogonal Model, individual items were treated as potential indicators for a given dimension if prior studies had produced an average loading of .25 (absolute) or greater and that loading was at least twice as large as the standard deviation across studies (cf., Table 1). The .25 criterion was lenient relative to most factor analytic guidelines, but it has been useful in prior research (Vickers, Conway, & Hervig, 1990). The requirement that the loading be twice as large as the estimated standard deviation can be regarded as an approximate jackknife test for the significance of the loading (Efron, 1981).

(d) Three-Dimensional Model: Many CES-D items had substantial loadings on both the depressed mood and functional disruption dimensions in Table 1. In contrast, items defining the third and fourth dimensions tended to have large loadings only on those dimensions. This pattern could indicate that too many dimensions have been extracted and rotated, so that a single underlying dimension has been split in two. This explanation seemed plausible given that rules regarding how many factors to extract in exploratory factor analyses are imprecise and that the common practice of extracting all factors with eigenvalues greater than 1.00 may extract too many factors (Zwick & Velicer, 1986). Therefore, a fourth measurement model was constructed to test the overfactoring hypothesis by combining the first and second dimensions of the Hertzog model into a single dimension. The model specified that the dimensions were correlated, because the three dimensions were believed to represent a single higher-order construct of depression. This decision regarding correlations between dimensions also facilitated comparisons to the Hertzog and Sample Weights models and provided a direct statistical test for correlations between dimensions rather than making an assumption that these correlations were orthogonal.

These four additional alternative models and the Hertzog model were compared by fitting each model to data generated by a large sample of U.S. Navy recruits. Additional exploratory models were derived on the basis of modification indices to better understand the differences between the focal models that best fitted the data.

Method

Sample

The sample (n = 2,746) was comprised of male volunteers participating in a series of studies pertaining to psychological predictors of health in basic training. The typical participant was 19.53 (SD = 2.67) years of age. Most participants had a high school diploma (88.0%) or Graduate Equivalency Degree (6.5%), but a small proportion had not completed high school

(5.5%). The primary ethnic groups were Caucasian (68.4%), Afro-American (14.9%), and Hispanic-American (8.1%) with other ethnic groups accounting for the remainder (7.4%).

Measures

The 20-item CES-D scale developed to measure depression in the general population (Radloff, 1977) was administered to recruits on the first day of basic training. The scale was administered verbally to the participants to avoid potential problems with reading skills and to ensure that the inventory was completed within time limits set for the data collection by the participant's training schedule. Respondents marked response options on an optical scanning sheet corresponding to the following scale:

1 = Rarely or none of the time (Less than 1 day)

2 = Some or a little of the time (1 - 2 days)

3 = Occasionally or a moderate amount of time (3 - 4 days)

4 = Most or all of the time (5 - 7 days)

Each question was read twice, and participants were encouraged to ask at that time for explanation or repetition of any item that was not clear to them. After all 20 items had been read, participants were given the opportunity to ask for repetition of any item they missed.

Analysis Procedures

An initial data screen eliminated cases with missing data or zero variance in responses to the CES-D items. Complete data was needed to ensure that estimates of covariances were mathematically consistent. The variability requirement was imposed on the assumption that failure to discriminate between items was reason to believe the person providing the data was inattentive or unwilling to participate (despite being told that participation was voluntary). While it is conceivable that a person could truly have zero variance, this situation was judged more likely to represent people who were providing questionable data. The variability requirement had only a slight impact on the overall sample, as only 0.5% (13 of 2759) of the recruits who completed the instrument had zero variance.

Confirmatory factor analyses were conducted by applying Joreskog and Sorbom's (1981) LISREL VI program to estimate parameter values and goodness-of-fit for the models described

in the introduction. The competing models were evaluated by standard chi-square evaluations and three goodness-of-fit indicators. Goodness-of-fit indicators were:

(a) The root mean square residual (RMS), i.e., the average difference between the estimated and observed covariances.

(b) The Tucker-Lewis Index (TLI; Tucker & Lewis, 1973) which is based on the reduction in the ratio of chi-square to degrees of freedom. TLI was chosen over alternative measures, because recent evidence indicates it is less influenced by sample size effects than are other measures (Marsh, Balla, & MacDonald, 1988).

(c) Parsimony indices (PIs) were computed from the TLI to take into account differences in the degrees of freedom for the competing models. This approach to model comparisons can be justified philosophically (Mulaik et al., 1989) and on the basis that parsimonious models produce smaller sampling errors for parameter estimates (Bentler & Mooijaart, 1989).

Results

The goodness-of fit results for the five models are given in Table 2 with the models ordered from the most constrained to the least constrained. The table shows that the Hertzog model accounted for 82.0% of the covariation between the CES-D items as indicated by the reduction in chi-square relative to the null model. The results in the table also show clearly that the Hertzog model did not fit the data as well as any model involving estimation of factor loadings from the data being analyzed. However, the fact that the PI for the Hertzog model was substantially higher than that for any other model indicates that the smaller chi-square values for the alternative models can be attributed largely to the greater opportunity to capitalize on chance when estimating parameter values from the data.

Table 2
Summary of Goodness-of-Fit Assessments

			Goodness-of-Fit Indices:		
Model	df	Chi-Square	RMS	TLI	PI
Null	190	14770.52	.353		
Hertzog	190	2660.94	.900	.831	.831
Unidimensional	170	2286.26	.054	.838	.750
Three-Dimensional	167	1535.51	.044	.893	.785
Sample Weights	164	1326.30	.041	.908	.783
Revised Sample	157	1147.49	.038	.918	.759

Two other important findings were evident in the results for the Unidimensional, 3-Dimensional, and Sample Weights models. First, a unidimensional model did not fit the data as well as either of the more complex alternatives. Second, the PI for the 3-dimensional model was slightly larger than that for the 4-dimensional sample weights model despite the statistical significance of the difference in fit for the two models (chi-square = 209.21, 3 df, $p < .001$).

Factor loadings estimated in the present analyses for the 3- and 4-dimensional models are given in Table 3. The shift from four dimensions to three dimensions had somewhat more effect on the weights for the Functional Disruption items than on the weights for the Depressed Mood items. Comparing the weights assigned in the four-dimensional solution to those in the three

Table 3
Maximum Likelihood Factor Loadings for the 3-Dimensional and Sample Weights Models

	3-Dimensional Solution			4-Dimensional Solution			
	1	2	3	1	2	3	4
1 Bothered by things	.567					.610	
2 Poor appetite	.413					.443	
3 Have the blues	.816			.819			
4 Feel as good as ever		.347			.346		
5 Can't keep mind on things	.589					.624	
6 Feel depressed	.906			.921			
7 Everything is an effort	.243					.264	
8 Hopeful about the future		.470			.469		
9 Feel like a failure	.511			.505			
10 Fearful	.653			.648			
11 Sleep is disrupted	.475					.527	
12 Feel happy		.640			.641		
13 Talk less than usual	.311					.329	
14 Feel lonely	.770			.782			
15 People are unfriendly			.585				.592
16 Enjoy life		.701			.701		
17 Feel like crying	.414			.418			
18 Sad	.839			.856			
19 People dislike me			.738				.730
20 Cannot get going	.562					.620	

NOTE: Factors are: 1 = Depressed Mood, 2 = Positive Affect, 3 = Functional Disruption, and 4 = Interpersonal Problems.

dimensional solution, the weights for items which defined the Functional Disruption dimension were between .018 and .058 smaller in the three-dimensional solution. The comparable figure was .003 to .017 for the Depressed Mood items. The most striking fact about these changes, however, was that the changes were small relative to the overall magnitude of the weights. Collapsing depressed mood and affective disruption to define the 3-dimensional model, therefore, produced a combined dimension with item loadings very similar to those obtained for the separate dimensions because these dimensions were highly correlated (r = .870).

Latent Trait Correlations. With one exception, the correlations between latent traits indicated less than 45% overlap in the variance of any two traits (Table 4). The exception was the correlation of .870 between Depressed Mood and Functional Disruption in the 4-dimensional model; although the absolute magnitude of this correlation is only 30% larger than the next largest correlation in the table, the proportion of overlap in variance for the two dimensions is 69% higher. Descriptively, therefore, the 4-dimensional solution produced much higher maximum overlap between latent traits than did the 3-dimensional model.

Table 4
Estimated Latent Trait Correlations

Latent Dimension:	1	2	3	1	2	3	4
1	1.000			1.000			
2	-.649	1.000		-.641	1.000		
3	.624	-.394	1.000	.870	-.604	1.000	
4				.593	-.396	.670	1.000

NOTE: Dimensions correspond to those in Table 3. For the 3-dimensional model, factors were: (1) Depression/ Disruption, 2 = Positive Affect, and 3 = Interpersonal Problems. For the 4-dimensional model, factors were: 1 = Depressed Mood, 2 = Positive Affect, 3 = Functional Disruption, and 4 = Interpersonal Problems.

Post-Hoc Model Modification. Modification indices (MIs) indicate the improvement in fit of the model that would be achieved by relaxing the constraints on a given parameter in the model (Joreskog & Sorbom, 1981). In the Hertzog model, each item had three MIs, one for each of the dimensions for which its loading was constrained to be zero. Examination of these indices suggested two primary points of discrepancy between the Hertzog model and the structure of the present data. For the effort item (#7), these MIs ranged from 224.58 to 467.13; comparable values for the happiness item (#12) ranged from 77.78 to 130.74 with three of the four values greater than 102.96. Fixing the variance on the functional disruption dimension at 1.00 also was a source of substantial misfit (MI = 113.24) as was fixing the correlation between Depressed Mood and Positive Affect at -.816 (MI = 116.81). The largest of the remaining 57 MIs (18 items, each with 3 constrained loadings plus 3 fixed variances) was 67.08, so there was a substantial discontinuity in the distribution of MIs between 77.78 and 102.96.

Comparison of the Sample Weights Model parameter estimates to the Hertzog model parameter estimates indicated that the Hertzog model weights were too high for effort (.783 versus .264) and happiness (.874 versus .641). When the Hertzog model was modified by freeing the constraint on the loading for effort, the improvement in fit was substantial (chi-square = 2142.18, 189 df, AGFI = .913; TLI = .865; PI = .861). At this point, the MIs for happy still were large (106.60 to 145.52), so the constraint on this parameter was relaxed with a further improvement in fit of the model to the data (chi-square = 2001.73, 188 df, AGFI = .925; TLI = .874; PI = .865). The constraint on the correlation between Depressed Mood and Positive Affect still had a large MI, so this constraint also was removed with a resulting improvement in fit (chi-square = 1908.11, 187 df, AGFI = .929; TLI = .880; PI = .866). The estimated parameter values obtained with the last of these three models were .262 for effort, .640 for happiness, and -.693 for the correlation between Depressed Mood and Positive Affect. Modifying these three constraints accounted for 56% of the original chi-square difference between the original Hertzog model and the model provided by estimating sample specific weights and sample specific correlations. The parsimony index for any of these three models was greater than that for the best model in Table 2.

Discussion

The development of a single measurement model for the CES-D which is robust across populations and situations is a realistic possibility. The implication of this finding in the present context is that measurement models and validity data developed for the CES-D scale in civilian populations can be generalized to military populations. This fundamental conclusion was supported by the goodness-of-fit of the Hertzog model to the recruit data. Naturally, the fit of this model to the data was not as good as the fit obtained when parameter estimates were made to maximize the fit between the model and the data, but the improvement in fit obtained with the sample-specific models considered here was never large enough to improve the parsimony index.

The results also illustrated the potential value of employing confirmatory factor analyses to test for generalizability of civilian measurement models to military populations. The bulk of the difference in fit between the original Hertzog et al. (1990) model and the sample-specific 4-dimensional model was localized in two factor loadings and one factor correlation. Given the differences between the populations and situations considered in the present study and those considered by Hertzog et al. (1990), this finding may represent the influence of the specific situation on reports of depressive symptoms. If so, large discrepancies between a standardized measurement model and the results obtained in a given data set may be an efficient method of identifying items which have taken on different meanings as a result of contextual factors. One application of a general model for the CES-D, therefore, could be as a frame of reference to diagnose such effects as a means of guarding against distortions of depression estimates. This potential application would be possible only if changes in structure typically are limited to a small proportion of the total parameter values under consideration as they were in this case.

The available evidence is insufficient for determining whether the observed misfit between the recruit data and the Hertzog et al. (1990) model really represents failure of a civilian model to generalize to the recruit population. The Hertzog et al. (1990) weights are affected by sampling variability in the estimation of the covariances and were not based on a large representative sample drawn from a general United States or North American population. Better estimates of the population parameter values would be obtained from large, representative samples of the general population. When these results are compared to those in the recruit population, differences between the two could be the result of sampling variation relative to the

population covariance matrix in either of the samples. A study of the CES-D structure in a large, representative sample of the United States/North American population would be useful to provide an appropriate frame of reference for interpreting the differences observed in the present study. A comparison of CES-D measurement models obtained from data acquired using the typical written format with the model obtained using the verbal presentation method employed in this study would test another possible explanation for the observed differences.

Any extension of the present work should give attention to the possibility that the CES-D should be represented by a 3-dimensional model or a 4-dimensional model. In the present data, these two models were closely comparable in terms of their fit to the data, but this comparability is contingent on the strong correlation between the affective and functional disruption dimensions ($r = .87$). This correlation was higher in the present study than in either the Annville ($r = .83$) or Victoria ($r = .71$) samples in the Hertzog et al. (1990) study. If the present results are closer to population values than those in the Hertzog et al. (1990) study, it could be argued reasonably that these two dimensions should be combined into a single dimension.

The results of this study indicate that measurement models for the CES-D can be generalized from the civilian population to military populations even when the military population is facing what may be extreme disruption in their social matrix. The results suggest that some minor changes in the measurement structure may occur as a result of situational factors, but it is not possible to be certain at this time that the apparent differences are not the product of chance variations. The sensitivity of the confirmatory factor analysis procedures to these minor differences suggests that this analytic approach can be a useful method of identifying such differences for careful evaluation whenever civilian models are generalized to military populations.

The study findings also raised some issues for future consideration in research with the CES-D which are not limited to studies of military populations. Further study to evaluate this instrument should give attention to the appropriate level of aggregation. Should the four components of depression be treated separately or should they be regarded as manifestations of a single syndrome? The presence of reliably defined subcomponents of depression does not resolve this issue. Neither is it certain that the common clinical practice of treating depression as a single general syndrome provides the best approach to achieving insight into this illness.

It can be argued that the appropriate model for the present data would be a general second-order factor that represents overall depression as was done by Hertzog et al. (1990) in some of their analyses. The correlations between the CES-D dimensions leave substantial room for doubt about whether to make distinctions between the various dimensions or to combine these assessments into a second-order factor of general depression.

One useful line of inquiry pertaining to the appropriate level of aggregation would be studies to determine whether reliable patterns of discriminant validity can be identified for specific CES-D dimensions. If no reliable patterns can be established, no purpose would be served in retaining a more complex model. Prior findings that positive and negative affect are largely independent empirically and are related to distinct elements of personality (e.g., Costa & McCrae, 1980; Diener & Emmons, 1984; Emmons & Diener, 1985; Tellegen, 1982; Watson, Clark, & Tellegen, 1988) provides some reason to believe that discriminant validity can be established relative to some criteria, but it is an open question whether alternative criteria such as biological markers for depression will show a similar pattern. The present models provide context for such research by defining alternative representations of primary level factors and by pointing up the strong correlations between measures of such factors. This latter point implies that large samples will be needed to provide statistically powerful tests of hypotheses pertaining to discriminant validity (Cohen, 1969).

The present findings indicate that the measurement structure of the CES-D is not dramatically affected by exposure to a novel, demanding environment. To the degree that this finding generalizes to the special demands imposed by other operational and training settings, the CES-D can be used to measure depression in military populations without concern that those situations have a major impact on the measurements. The next logical step for improving the measurement of depression with the CES-D in both military and civilian populations is the development of parameter estimates for both 3- and 4-dimensional models in a large representative cross-section of the population. These estimates then could be used as a gold standard against which to evaluate structures in specific populations and situations, including recruits and other subpopulations of interest in the military. These estimates also could be used to fix the measurement models in studies testing the distinctiveness of patterns of association of specific depression components with potential causes and consequences of depression. If

distinctive patterns cannot be demonstrated for the various dimensions of depression, a unidimensional representation of depression would be preferred to either of the models considered here on the basis of parsimony arguments.

References

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. Psychological Bulletin, 106, 315-317.

Bentler, P.M., & Woodward, A. J. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. Psychometrika, 45, 249-267.

Berkman, L. F., Berkman, C. S., Kasl, S., Freeman, D. H., Leo, L., Ostfeld, A. M., Cornoni-Huntley, J., & Brody, J. A. (1986). Depressive symptoms in relation to physical health and functioning in the elderly. American Journal of Epidemiology, 124, 372-388.

Blalock, H. M., Jr. (1982). Conceptualization and measurement in the social sciences. Beverly Hills, CA: Sage.

Bourne, P. G. (1967). Some observations on the psychosocial phenomena seen in basic training. Psychiatry, 30, 187-196.

Clark, V. S., Aneshensel, C. S., Frerichs, R. R., & Morgan, T. M. (1981). Analysis of effects of sex and age in response to items on the CES-D scale. Psychiatry Research, 5, 171-181.

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. NY: Academic Press.

Costa, P. T., Jr., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. Journal of Personality and Social Psychology, 38, 668-678.

Datel, W. E., Engle, E. O., & Barba, M. A. (1966). Affect levels in a company of basic trainees. Psychological Reports, 19, 903-909.

Devins, G. M., Orme, C. M., Costello, C. G., Binik, Y. M., Frizzell, B., Stam, H. J,. & Pullin, B. M. (1988). Measuring depressive symptoms in illness populations: Psychometric properties of the Center for Epidemiologic Studies Depression (CES-D) scale. Psychology and Health, 2, 139-156.

Diener, E. & Emmons, R. A. (1984). The independence of positive and negative affect. Journal of Personality and Social Psychology, 47, 1105-1117.

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap, and other methods. Biometrika, 68, 589-599.

Emmons, R. A., & Diener, E. (1985). Personality correlates of subjective well-being. Personality and Social Psychology Bulletin, 11, 89-97.

Ensel, W. M. (1986). Measuring depression: The CES-D scale. In N. Lin, A. Dean & W. Ensel (Eds.), Social support, life events, and depression. NY: Academic Press, pp. 51-70.

Fordyce, M. W. (1988). A review of research on happiness measures: A sixty second index of happiness and mental health. Social Indicators Research, 20, 355-381.

Gorsuch, R. (1973). Factor analysis. Philadelphia: Saunders.

Hertzog, C., Van Alstine, J., Usala, P. D., Hultsch, D. F., & Dixon, R. (1990). Measurement properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations. Psychological Assessment, 2, 64-72.

Joreskog, K. G., & Sorbom, D. (1981). LISREL: Analysis of linear structural relationships by the method of maximum likelihood. Chicago: International Educational Services.

Kuo, W. H. (1984). Prevalence of depression among Asian-Americans. Journal of Nervous and Mental Disease, 172, 449-457.

LaRocco, J. M., Ryman, D. H., & Biersner, R. J. (1977). Life history and mood as predictors of adjustment in Navy recruit training. Journal of Community Psychology, 5, 46-51.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). Profile of Mood States manual. San Diego, CA: Educational and Industrial Testing Service.

Meehl, P.E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. Psychological Inquiry, 1, 108-141.

Meehl, P.E. (1991). Why summaries of research on psychological theories are often uninterpretable. In R. E. Snow & D.E. Wiley (Eds.), Improving inquiry in social science (pp. 13-60). Hillsdale, NJ: Erlbaum.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stillwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement, 1, 385-401.

Regier, D. A., Boyd, J. H., Burke, Jr., J. D., Rae, D. S., Myers, J. K., Kramer, M., Robins, L. N., George, L. K., Karmo, M., & Locke, B. Z. (1988). One-month prevalence of mental disorders in the United States. Archives of General Psychiatry, 45, 977-986.

Roberts, R. E. (1980). Reliability of the CES-D scale in different ethnic contexts. Psychiatry Research, 2, 125-134.

Ross, C. E., & Mirowsky, J. (1984). Components of depressed mood in married men and women. American Journal of Epidemiology, 119, 997-1004.

Ryman, D. H., Biersner, R. J., & LaRocco, J. A. (1974). Reliabilities and validities of the mood questionnaire. Psychological Reports, 35, 479-484.

Tellegen, A. (1982). Brief manual for the Differential Personality Questionnaire. Unpublished manuscript, University of Minnesota, Minneapolis, MN.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Vickers, R. R., Jr., Conway, T. L., & Hervig, L. K. (1990). Demonstration of replicable dimensions of health behaviors. Preventive Medicine, 19, 377-401.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063-1070.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. Psychological Bulletin, 98, 219-235.

Wolff, C. T., Hofer, M. A., & Mason, J. W. (1964). Relationship between psychological defenses and mean urinary 17-hydroxycorticosteroid excretion rates. II. Methodologic and theoretical considerations. Psychosomatic Medicine, 26, 592-609.

Zurcher, L. A., Jr. (1968). The naval recruit training center: A study of role assimilation in a total institution. Sociological Inquiry, 31, 85-98.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99, 432-442.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
February 1992

3. REPORT TYPE AND DATE COVERED
Interim Oct - Dec 1991

4. TITLE AND SUBTITLE
Confirmatory Factor Analysis of the Center for Epidemiological Studies Depression Scale (CES-D in Military Recruits

5. FUNDING NUMBERS
Program Element:
Work Unit Number:
MR04101.00A-6004
ONR.WR.24030

6. AUTHOR(S)
Ross R. Vickers, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION
Report No. 92-29

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center
Building 1, Tower 2
Bethesda, MD 20889-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Psychological assessment procedures often are generalized from civilian populations to military personnel. Generalization can be misleading if the demanding training and operational conditions in the military produce the appearance of score elevations because of normal reactions to specific situational factors that may be more common in the military than in civilian populations (e.g., sustained operations). Such situation-specific effects would be expected to change the pattern of covariation between the affected indicators and others comprising the measurement model, so differences in measurement structure in the military and civilian populations could be one indicator that generalization was questionable. The present study applied confirmatory factor analyses of responses to the Center for Epidemiological Studies depression scale (CES-D) obtained from male U.S. Navy recruits (N=2,746) one week after arrival at basic training to test for such effects. This test was chosen because depression is an important psychological index of well-being and because basic training involves situational changes that could increase the reporting of apparent depressive symptoms. Using parsimony criteria to choose between several competing models, the best model tested involved modifications of only 3 of 26 parameter values comprising the civilian model. The CES-D measurement model developed and validated in civilian populations can be generalized to military populations with some confidence, but confirmatory analyses such as the present ones can use misfit between the data and a confirmatory measurement model to identify possible biases arising from such generalization.

14. SUBJECT TERMS
depression
confirmatory factor analysis
military recruits
Center for Epidemiological Studies Depression scale

15. NUMBER OF PAGES
21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited